基于似物目标的快速行人检测算法 *

刘 倩[1,2]，李 策[1]，杨 峰[1]

(1. 中国矿业大学(北京) 机电与信息工程学院，北京 100083; 2. 宁夏大学 信息工程学院，银川 750001)

摘 要：针对传统滑动窗口行人检测速度慢、无效窗口数量大和识别率低的问题，提出基于似物目标的快速行人检测算法。该算法首先使用改进的二值化赋范梯度算法生成目标候选集，再利用提出的 NDOG 特征描述子计算目标特征。通过在 INRIA 和 Caltech-USA 两个行人数据库分析实验，证明该算法可有效减少待检测窗口数量，提高特征表示能力，且计算时间显著降低，使用线性支持向量机作为分类器可获得较高的行人检测率。

关键词：似物目标检测；二值化赋范梯度；方向梯度直方图；极限学习机

中图分类号：TP391.4　　doi: 10.3969/j.issn.1001-3695.2018.01.0115

Rapid pedestrian detection based on generic object generation

Liu Qian[1, 2], Li Ce[1], Yang Feng[1]

(1. *School of Mechanical Electronic & Information Engineering*, *China University of Mining & Technology*, *Beijing* 10083, *China*; 2. *College of Information Engineering Ningxia University*, *YinChuan* 750001, *China*)

Abstract: Detection of pedestrians on the image using a sliding window mode results in a huge amount of invalid windows and inefficiency. To circumvent these problems for pedestrian detection, this paper proposed an approach, based on generic object generation, for rapid detection of pedestrian. Firstly, the approach generated the generic object proposals by the Improved Binarized Normed Gradients, and then calculated the character of proposals based on our NDOG feature descriptor. Experimental results over the INRIA and Caltech-USA pedestrian datasets show that our approach reduces the number of proposals effectively, and improves the efficiency and the detection accuracy significantly along with a linear SVM classifier.

Key words: generic object proposal generation; BING; HOG; ELM

0　引言

在科技水平飞速发展、计算机运算能力不断提高的背景下，如何使计算机更好的满足人们生产生活的实际需要是计算机视觉领域的实现目标。其中，模拟人眼视觉系统，通过对摄像头拍摄的画面进行识别和分析，提取目标对象是该领域工作者的研究热点，而自然场景下的行人识别则是该领域的重要组成部分。

基于梯度直方图（histogram of oriented gradient， HOG）[1]的图像描述方法是当前人体模型建模的主要方法之一。该算法通过对图像中每个像素点的梯度幅值和梯度方向做统计的方法来描述窗口中像素点梯度分布情况。与其他常用的特征表示方法相比，HOG 可较好的表征目标对象的轮廓线条，提取的特征鲁棒性强，算法简单易懂，代码量少。

为提高 HOG 特征的表征能力，此后产生了很多基于 HOG 的改进算法[2-5]，。但这些方法都无法实现快速的行人检测，且都忽略了相邻 block 之间的空间关系，而这个因素对梯度直方图检测方法来说是至关重要的。

在搜索策略方面，行人检测系统[6, 7]通常先计算滑动窗口（sliding window）中的特征再判断该窗口是否包含行人特征。而在实际计算过程中，一副图像对应的滑动窗口数量是巨大的，并且其绝大部分是无效窗口，即不包含目标对象或只包含目标对象的一部分，因此如何减少待检测窗口的数量是提高目标识别效率的关键。似物性目标检测方法[8, 9]根据人眼感知对象的原理[10]，可快速定位图像中前景目标，为在行人识别的过程中降低检测窗口数量、提高识别效率提供了有效的解决方法。

在似物目标检测算法中[11-13]，二值化赋范梯度（binarized normed gradients，BING）检测算法利用二级支持向量机（support vector machine，SVM）[14]对待测目标的梯度图像进行检测。算法中对梯度图像和一级支持向量机模板做二值化处理，通过二进制运算快速实现梯度图像和模板的内积计算。与其他似物目标检测算法相比，BING 算法仅需要 3 毫秒即可完成目标检索。因此，使用 BING 对图像中似物目标快速检测，为降低检测窗口数量，实现快速行人检测提供了可能。但 BING 算法仍有问

收稿日期：2018-01-20；**修回日期**：2018-03-08　**基金项目**：贵州省科技计划资助项目（黔科合 GZ 字[2015]3020）；国家自然科学基金资助项目（61751215）

作者简介：刘倩（1981-），女，副教授，博士，主要研究方向为计算机应用、多媒体处理（lq_nxdx@163.com）；李策（1988-），女，讲师，主要研究方向为计算机视觉；杨峰（1968-），男，教授，博士，主要研究方向为计算机图形图像处理.

题亟待解决，如文献[15]提出无论一级支持向量机的模板如何调整，对其算法检测率的影响均不明显。但作者没有解释造成这一问题的原因，也没有提出相应的办法予以解决。

本文通过对上述问题的认真分析，提出了基于似物目标的快速行人检测算法。首先，利用改进的 BING 算法快速识别图像中似物目标，然后结合改进的 HOG 方法——方向梯度邻域描述子方法（neighborhood descriptor of oriented gradient，NDOG）进行行人检测。使用国际上常用的公开行人数据集测试，并与原算法进行比较。实验表明，本文提出的算法显著降低候选窗口数量，节省计算时间，并达到较高的检测率。

1　改进的似物性检测 BING 算法

1.1　BING 算法

在 2014 的 IEEE 国际计算机视觉与模式识别会议（IEEE Conference on Computer Vision and Pattern Recognition，简称 CVPR）程明明等人发表的《Binarized Normed gradients for objectness estimation at 300fps》中提出了 BING 算法。此算法在具有较高检测率的同时其运算速度远远超过其他同类算法。在 BING 算法中，高效运行的关键是利用二进制移位运算和或运算实现测试图像和一级分类模板的内积运算得到第一阶段的似物性评估分值。尽管这一部分实现了快速内积运算，但也是文献[15]对 BING 算法提出的问题所在。即无论如何调整其分类模板 $w \in R^{64}$ 中权值，其窗口检测率与原算法的检测率没有明显差异。

本文在对 BING 分析中发现，该算法在识别过程第一阶段仅包含一个二值线性分类器 $w \in R^{64}$，该分类器是基于 Liblinear[16]的 squared hinge loss（或简称 L2 loss）支持向量回归（support vector regression，SVR）方法计算得出。BING 实际上是利用模板 W 线性拟合其他训练样本。为了量化 BING 在第一阶段的识别性能，在相同训练样本集下，本文使用同样基于 Liblinear 的 L2 loss SVM 算法对测试样本进行分类识别。图 1 显示第一阶段分类器 W 对 PASCAL VOC2007[17]测试样本的似物目标检测结果。从图中可知，BING 算法的准确率和灵敏度分别是 73.67%和 100%，但特异度只有 0.0001%。这一实验结果说明所有测试样本均被判断为正样本。由此可以得出结论，BING 算法中二值线性分类器 $w \in R^{64}$ 不具备识别样本的能力。

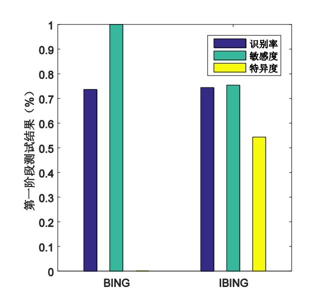


图 1 BING 算法第一阶段的似物目标检测性能指标

极限学习机[18~20]（extreme learning machine，ELM）作为求解单隐层前馈神经网络的算法，其隐藏输入节点权值随机初始化，输出权值根据 Moore-Penrose 广义逆[21]方法计算得出。该算法执行过程中不需要调整网络的输入权值以及隐元的偏置，并且产生唯一最优解，因而学习速度快且泛化性能好，其运算速度高于 SVM 算法。ELM 也同 SVM 算法一样，可以通过优化推导的方法得出与 Moore-Penrose 广义逆相同的输出权值。更重要的是，ELM 算法中输入节点权值也可按照 BING 算法中的方法实现二值化。这为将 BING 算法中的 SVM 算法替换为 ELM 提供了可能。从算法实现原理角度分析，ELM 隐层中由多个节点组成，样本需要与所有隐层节点做内积运算，其共同运算的结果即为算法输出值。若将 BING 算法中模板 w 替换为一组模板 $w_1,\ldots,w_L$，可明显提高对样本的识别能力。图 1 显示了将 BING 算法中第一阶段的 SVM 算替换为 ELM 方法后各项分类评价指标。图中改进算法的识别率是 74.43%，较原算法相比识别率提高了 0.76%。虽然灵敏度从指标上观察下降了 24.58%，但特异度提高了 54.35%。特异度的大幅提升说明在 BING 算法中利用 ELM 代替 SVM 的方法可以提升对负样本的识别能力，大幅减少 BING 算法在第一阶段计算得到的候选窗口的数量。

1.2　BING 的改进算法

为了更清楚地理解 BING 算法的改进方法（improved BING，IBING），本文首先对 BING 实现步骤做简要介绍。假设有 N 个训练样本 (x_i,t_i)，其中 $x_i=[x_{i1},x_{i2},\ldots,x_{in}]^T \in R^n$，$t_i=[t_{i1},t_{i2},\ldots,t_{im}]^T \in R^m$。在单隐层前馈神经网络中可表示为

$$H\eta = T \tag{1}$$

其中：η、T 和 H 分别表示输出权值、期望输出和隐层输出矩阵。H 表示为

$$H(w_1,\ldots,w_L,d_1,\ldots,d_L,x_1,\ldots,x_N)=\begin{bmatrix} H_1 \\ \vdots \\ H_N \end{bmatrix}$$

$$=\begin{bmatrix} G(w_1\cdot x_1+d_1) & \cdots & G(w_L\cdot x_1+d_L) \\ \vdots & \cdots & \vdots \\ G(w_1\cdot x_N+d_1) & \cdots & G(w_L\cdot x_N+d_L) \end{bmatrix}_{N\times L} \tag{2}$$

$w_i=[w_{i1},w_{i2},\ldots,w_{in}]^T$ 和 d_i 分别表示第 i（i=1，…，L）个隐节点的输入权值和偏差，$G(\bullet)$ 是激活函数。

在随机初始化 w_i 和 d_i 后，根据极小范数最小二乘解方法确定隐层输出矩阵

$$\hat{\eta} = H^{\dagger}T\,, \tag{3}$$

其中：$H^{\dagger}$ 是矩阵 H 的 Moore-Penrose 广义逆。由此计算得到 ELM 算法中输出权值 η，并利用式（1）判断样本类别。

在 IBING 算法中，将 BING 算法中 SVM 算法替换为 ELM 的关键问题就是如何实现 $w\cdot x$ 二值化。首先根据式（4）（5）对 w 和 x 做二值化处理

$$w=\sum_{j=1}^{N_{basis}}\beta_j a_j\,, \tag{4}$$

$$x=\sum_{j=1}^{N_{basis}}\alpha_j b_j \tag{5}$$

其中：N_{basis} 是偏差向量的个数，$\alpha_j,\beta_j \in R$ 是系数，$a_j,b_j \in \{-1,1\}^{N_{basis}}$ 是偏差向量并且 $a_j^+,b_j^+ \in \{0,1\}^{N_{basis}}$ 满足条件

$$a_j = a_j^+ - \overline{a_j^+}, \tag{6}$$

$$b_j = b_j^+ - \overline{b_j^+}。 \tag{7}$$

则 w 和 x 的内积可表示为

$$\begin{aligned}\langle w,x\rangle &= \sum_{i=1}^{N_{basis}} \beta_i \langle a_i, \sum_{j=1}^{N_{basis}} \alpha_j b_j\rangle \\ &= \sum_{i=1}^{N_{basis}} \beta_i \sum_{j=1}^{N_{basis}} \alpha_j \langle a_i, b_j\rangle\end{aligned} \tag{8}$$

其中:

$$\begin{aligned}\langle a_i,b_j\rangle &= \langle a_i^+,b_j\rangle - \langle \overline{a_i^+},b_j\rangle \\ &= 2^2\langle b_j^+,a_i^+\rangle - 2^1[|a_i^+| + |b_j^+|] + 2^6\end{aligned} \tag{9}$$

因此式（8）可表示为

$$\langle w,x\rangle = \sum_{i=1}^{N_a}\beta_i \sum_{j=1}^{N_a}\alpha_j S_{ij} \tag{10}$$

其中：$S_{ij} = [2^2\langle b_j^+,a_i^+\rangle - 2^1[|a_i^+| + |b_j^+|] + 2^6]$ 可通过二进制移位运算和或运算实现。由此，实现了 BING 算法中 SVM 替换为 ELM。

2　NDOG 特征描述子

HOG 首先对灰度图像整体进行边缘检测，利用边缘算子计算各个像素点的梯度幅值和梯度方向，然后统计各个 cell 内相同方向通道的梯度累计直方图，最后由多个 cell 构成一个 block，block 内所有 cell 的梯度直方图串联得到该 block 的 HOG 特征直方图。HOG 算法允许相邻 block 间的存在 cell 重叠以此提高数据关联性。按照 Dalal 在文献中提出的算法，在大小为 64×128 的滑动检测窗口中设定 cell 尺寸为 8×8 像素点组成，block 由 2×2 块 cell 组成，block 移动步长为 8 个像素单位构成的 HOG 描述子对图像中行人检测的漏检率是最低的。但 HOG 特征长度为 3780 维，每个 cell 平均被统计 4 次，描述信息里有四分之三的数据是重复的，冗余信息量很大。且该算法没有利用各个 cell 之间的局部邻域关系。

针对上述问题，结合相邻样本具有线性相关性的特点[22]，提出了方向梯度邻域描述子（NDOG）方法。该算法不需要重叠 block，仅将每个 block 与周围相邻的 8 个 block 的邻域关系通过权值向量表示，并将该向量串联到每个 block 方向直方图向量的尾部。

假设样本由 m×n（=N）个 block 组成，block $b^i \in R^{36}$（i=1，…，N）根据 HOG 算法得出。计算 b^i 与其相邻的 8 个 block 之间的权值系数 $W^i = [W_1^i, W_2^i, \ldots, W_8^i]$，并保证重构代价误差最小：

$$\varepsilon(W) = \sum_{i=1}^{N}\left\| b^i - \sum_{j=1}^{8} W_j^i b_j^i \right\|^2 \tag{11}$$

其中权值 W_j^i 表示 b^i 与第 j 个相邻 block 的相似度，式(11）满足约束条件 $\sum_{j=1}^{8} W_j^i = 1$。重构代价误差亦可写成

$$\begin{aligned}\varepsilon(W^i) &= \left\| b^i - \sum_{j=1}^{8} W_j^i b_j^i \right\|^2 \\ &= \left\| \sum_{j=1}^{8} W_j^i b^i - \sum_{j=1}^{8} W_j^i b_j^i \right\|^2\end{aligned} \tag{12}$$

定义 $B^i = [b^i, b^i, \ldots, b^i]_{36\times 8}$，$N^i = [b_1^i, b_2^i, \ldots, b_8^i]_{36\times 8}$，则

$$\begin{aligned}\varepsilon(W^i) &= \left\| B^i W^i - N^i W^i \right\|^2 \\ &= W^{i^T}(B^i - N^i)^T (B^i - N^i) W^i \\ &= W^{i^T} G^i W^i\end{aligned} \tag{13}$$

其中 $G_{jm}^i = (b^i - b_j^i)^T (b^i - b_m^i)$ 是对称半正定阵。根据拉格朗日乘子法在得到

$$L(W^i) = W^{i^T} G^i W^i + \lambda(W^{i^T}\vec{1} - 1) \tag{14}$$

计算式（14）中 W^i 的一阶偏导得

$$\frac{\partial L}{\partial W^i} = 2G^i W^i + \lambda\vec{1} = 0 \tag{15}$$

$$W^i = -\frac{\lambda}{2}(G^i)^{-1}\vec{1} = -\frac{\lambda}{2}\begin{bmatrix}\sum_{m=1}^{8}(G^i)^{-1}{}_{1m} \\ \vdots \\ \sum_{m=1}^{8}(G^i)^{-1}{}_{8m}\end{bmatrix}。 \tag{16}$$

根据条件 $\sum_{j=1}^{8} W_j^i = 1$ 计算得到每个 block 的邻域点权值向量

$$W_j^i = \frac{\sum_{m=1}^{8}\left(G^i\right)^{-1}_{jm}}{\sum_{p=1}^{8}\sum_{q=1}^{8}\left(G^i\right)^{-1}_{pq}}。 \tag{17}$$

在实际计算时，G^i 有可能是奇异矩阵，需要做正则化处理：

$$G^i = G^i + \alpha \cdot I。 \tag{18}$$

其中 α 是正则化参数，I 是 8×8 单位矩阵。

从式（18）可以看出相邻 block 见的欧几里得距离越近，其相似度越高，对应的权值越大。图 2 显示了相邻 block 的相似度和权值的关系。

HOG 算法原理是利用图像中目标和背景间梯度幅值的差异以及梯度方向的连续性建立目标特征。图 3（b）显示了图 3（a）中行人的 HOG 特征。从图 3（b）中可以看出行人轮廓处的 block 方向具有连续性和一致性，但在 HOG 算法中没有量化 block 间的这一重要信息。

因此，本文提出的 NDOG 算法的目标之一就是量化 block 间的相似性，图 3（c）显示了图 3（b）中 block 之间的相似性。为了让读者获得更直观的感受，将权值小于均值的 W_j^i 设为 0。如图所示，行人轮廓处的 block 相似性高于图像中的其他部分。鉴于 block b^i 与其权值向量 W_j^i 一一对应，本文提出将权值向量串联到对应 block 的尾部，为保证对比标准化，需要对特征向量再使用二阶梯度范数做归一化处理，串联后组成的向量即为 NDOG 特征描述子。

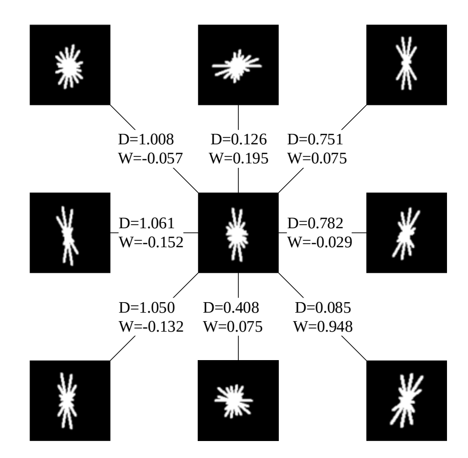
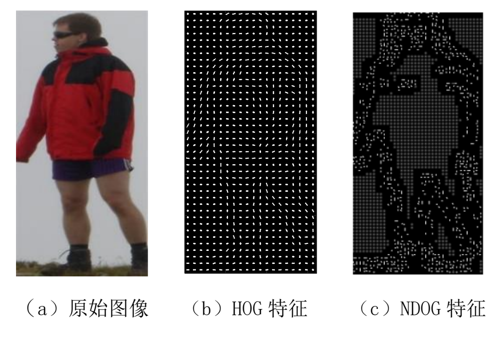


图 2 HOG 算法中 block 与相邻 8 个 block 相似度和权值关系示意图



（a）原始图像　　（b）HOG 特征　　（c）NDOG 特征

图 3 行人特征描述示意图

3　本文算法

本文是在基于似物性目标检测框架的基础上实现快速行人识别。首先使用本文提出的 IBING 算法得到似物目标候选集，然后计算似物性目标的 NDOG 特征并利用 SVM 进行进一步识别，具体算法如图 4 所示。

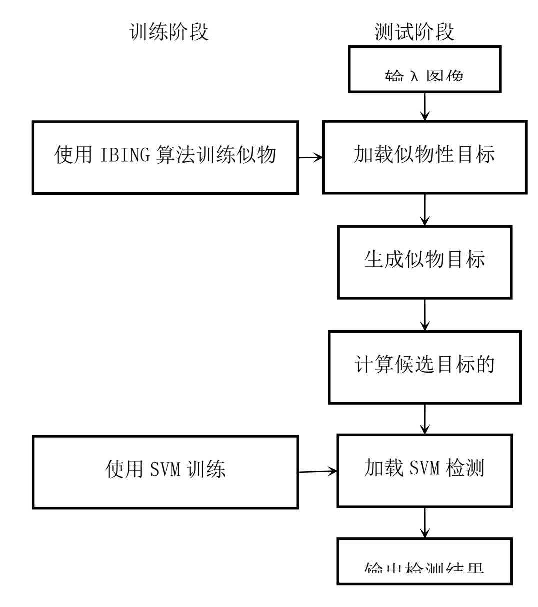


图 4 本文提出的算法流程

4　实验过程和结果分析

本文中的实验平台为 Intel(R) Core™ i5-4590 CPU @3.30 GHz 3.30 GHz，RAM 8.00 GB，64 bit operating system，MATLAB R2014b。为验证算法性能，本文使用 INRIA[1]和 Caltech-USA[23, 24]数据库作为实验用样本集。这两个数据库是行人识别邻域中公认的标准数据集，因此使用它们对提出算法进行性能分析和比较具有实际意义。此外，本文使用 LIBSVM[14]算法实现 SVM 过程的训练和分类。

4.1　候选窗口数量对比

选用 INRIA 数据库作为实验对象，分别使用滑动窗口、BING 算法和 IBING 算法计算 INRIA 每个测试图像候选窗口的数量。如图所示，在每个 INRIA 的测试图像上使用滑动窗口的方法因为不对窗口内图像做任何识别操作，所以候选窗口数量最多，每幅测试图像产生的滑动窗口数量大约在 1.2 万~2 万个间。BING 算法生成的候选窗口数量次之，但每个测试样本的候选窗口数量仍保持在 1 万个左右。导致这一现象的原因正是 BING 算法在第一阶段将所有样本判断为正样本的结果。IBING 算法在隐层包含多个匹配模板，故对样本的识别性能较 BING 算法有明显提升。在 INRIA 数据库的每个测试样本中得到的候选窗口数量只有 0.15 万个左右，远远低于前两种方法得到的窗口数量。有效降低了后续行人识别的计算量，节省结算时间。

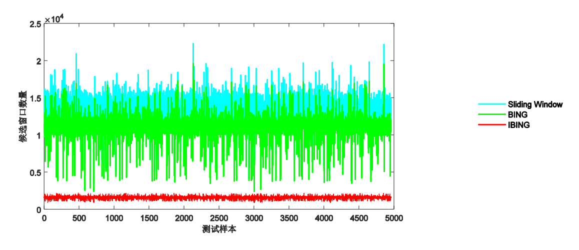


图 6 候选窗口数量对比图

4.2　行人特征对比

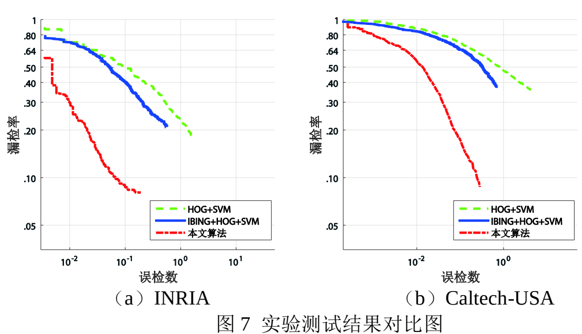
NDOG 是 HOG 的改进算法，其改进之处是在避免产生特征冗余信息。在文献[1]中，若定义滑动窗口的大小为 $M \times N$ 个像素点组成，block 由 $m \times m$ 个像素点组成，cell 由 $n \times n$ 个像素点构成，Block 在滑动窗口中移动步长为 k 个像素单位（ m 能被 k 整除），在一个 cell 内的像素点梯度方向分为 i 个梯度方向，则在滑动窗口中 HOG 描述子的长度 $=[(M/k-1)\times(N/k-1)]\times(m^2/n^2)\times i$ 。当 $M=64$，$N=128$，$m=16$，$n=8$，$k=8$，$i=9$ 时，HOG 描述子的长度是 3780 维。NDOG 算法在计算特征描述子时不需要重叠 block，只需要计算 cell 邻接点权值向量，并将向量（该向量的特征长度为 8）添加到该 cell 梯度直方图向量的尾部即可。因此，NDOG 描述子的长度=(M/n)×(N/n)×(i+8)=(64/8)×(128/8)×(9+8)=2176。与 HOG 特征相比，NDOG 特征向量长度明显缩短，不包含冗余信息，对行人特征的表示能力优于 HOG（图 3 所示）。

4.3　行人识别评价指标对比

实验对比了 HOG+SVM，IBING+HOG+SVM 和本文提出的 IBING+NDOG+SVM 算法在 INRIA 和 Caltech-USA 两种数据集的识别性能，评价指标本文采用 MR-FPPI（Miss Rate-False Positive Per Image）曲线[1]。实验结果如图 7 所示，IBING+HOG+SVM 的误检数量较原算法有明显降低，但漏检率也略有升高。这是因为 IBING 算法将正负样本的分界面向正样本方向移动，造成了在提高了负样本的检测能力的同时也将少量正样本划分到负样本集内。但是，使用本文提出的算法识别图像中的行人，MR-FPPI 性能要优于原算法和 IBING+HOG+SVM 算法。这是由于对候选窗口使用 NDOG 描述子提高了行人特征的表征能力，为 SVM 提供更加有效的样本以供训练和识别。

为了量化上述对比图的信息，表 1 显示了原算法和改进算法在 INRIA 数据库的具体的平均漏检率和识别速度。其中平均漏检率与图 7（a）表现一致，较原算法相比提高了 43%，并且

识别速度提高了 4 倍。



（a）INRIA （b）Caltech-USA

图 7 实验测试结果对比图

表 1 算法识别率和检测时间对比

算法名称	识别率	平均检测时间（秒）
HOG+SVM	70.35%	1.63
IBING+HOG+SVM	70.48%	0.28
文本算法	75.34%	0.46

5 结束语

在围绕如何实现快速行人检测这一问题时，本文提出使用似物目标检测算法选取候选目标再判断类别的方法达到快速识别行人的目的。通过对 BING 算法的改进，有效降低了候选窗口的数量，为实现后续快速行人识别工作提供了基础。本文提出的 NDOG 方法通过引入邻域权值信息，克服了 HOG 算法中信息冗余度高、特征表示不显著等问题。实验结果表明，本文提出的算法可大幅提升行人检测的实时性，并且识别性能有较明显的提高。下一步将研究如何结合视频时间信息，将同一行人的特征描述子建立时间坐标轴的描述信息，扩大本文方法在视频检测领域的应用范围。

参考文献：

[1] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005: 886-893.

[2] Walk S, Majer N, Schindler K, etc. New features and insights for pedestrian detection [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2010: 1030-1037.

[3] Watanabe T, Ito S, Yokoi K. Co-occurrence histograms of oriented gradients for human detection [J]. IPSJ Trans on Computer Vision and Applications, 2010 (2): 39-47.

[4] Felzenszwalb P F, Girshick R B, McAllester D. Cascade object detection with deformable part models [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2010: 2241-8.

[5] 刘琼，陈雯柏. 基于视觉选择性注意与 IHOG-LBP 特征组合的行人目标快速检测 [J]. 计算机应用研究, 2016, 33 (1): 281-285. (Liu Qiong, Chen Wenbai. Fast pedestrian detection method based on combinatory features IHOG-LBP and visual selective attention computation [J]. Application Research of Computers, 2016, 33 (1): 281-285.

[6] 曲永宇，刘清，郭建明，等. 基于 HOG 和颜色特征的行人检测 [J]. 武汉理工大学学报, 2011, 33 (4): 134-138. (Qu Yongyu, Liu Qing, Guo Jianming, *et al.* HOG and color based pedestrian detection [J]. Journal of Wuhan University of Technology, 2011, 33 (4): 134-138.)

[7] 疏坤，蒋建国，齐美彬，等. 基于改进的 HOG 与 Sobel-LBP 融合的快速行人检测 [J]. 合肥工业大学学报：自然科学版, 2017, 40 (7): 898-903. (Shu Kun, Jiang Jianguo, Qi Meibin, *et al.* Fast pedestrian detection based on combinatory feature of improved HOG and Sobel-LBP [J]. Journal of Hefei University of Technology (Natural Science, 2017, 40 (7): 898-903.)

[8] Alexe B, Deselaers T, Ferrari V. What is an object? [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010: 73-80.

[9] Hosang J, Benenson R, Dollár P, etc. What makes for effective detection proposals? [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2016, 38 (4): 814-830.

[10] Wolfe J M, Horowitz T S What attributes guide the deployment of visual attention and how do they do it? [J]. Nat Rev Neurosci, 2004, 5 (6): 495-501.

[11] Cheng M M, Zhang Z, Lin W Y, *et al.* BING: binarized normed gradients for objectness estimation at 300fps [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3286-3293.

[12] Zitnick C L, Dollár P. Edge Boxes: locating object proposals from edges [C]// Proc of the 13th European Conference on Computer Vision. 2014: 391-405.

[13] Qi Y, Song Y Z, Xiang T, *et al.* Making better use of edges via perceptual grouping [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1856-1865.

[14] Chang C C, Lin C J. LIBSVM: a library for support vector machines [J]. ACM Trans on Intelligent Systems and Technology, 2011, 2 (3): articleNo 27.

[15] Zhao Q, Liu Z, Yin B. Cracking BING and beyond [C]// Proc of British Machine Vision Conference. 2014.

[16] Fan R E, Chang K W, Hsieh C J, *et al.* LIBLINEAR: a library for large linear classification [J]. Journal of Machine Learning Research, 2008, 9 (6): 1871-1874.

[17] Everingham M, Van Gool L, Williams C K I, *et al.* The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88 (2): 303-338.

[18] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70 (1): 489-501.

[19] Huang G B. What are extreme learning machines?Filling the gap between frank rosenblatt's dream and John von Neumann's puzzle [J]. Cognitive Computation, 2015, 7 (3): 263-278.

[20] Huang G B, Wang D H, Lan Y. Extreme learning machines: a survey [J]. International Journal of Machine Learning and Cybernetics, 2011, 2 (2): 107-122.

[21] Banerjee K S. Generalized inverse of matrices and its applications by C. Radhakrishna Rao, Sujit Kumar Mitra [J]. Technometrics, 1973, 15 (1): 197.

[22] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290 (5500): 2323-2326.

[23] Dollar P, Wojek C, Schiele B, etc. Pedestrian detection: a benchmark [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2009: 304-311.

[24] Zhang S, Benenson R, Omran M, *et al.* How far are we from solving pedestrian detection? [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1259-1267.